SERVICE FOR PROVIDING SPEAKER VOICE METRICS

Inventor(s):

Thomas E. Creamer

Peeyush Jaiswal

Victor S. Moore

International Business Machines Corporation

IBM Docket No. BOC9-2003-0090

IBM Disclosure No. BOC8-2003-0121

SERVICE FOR PROVIDING SPEAKER VOICE METRICS

BACKGROUND

Field of the Invention

[0001]   The invention relates to communications systems and, more particularly, to conveying biometric information over a communications system.

Description of the Related Art

[0002]   When one caller places a telephone call to another caller, it may be difficult for one caller to determine how the other caller, or speaker, is feeling from an emotional perspective. Without the aid of visual images, oftentimes, a listener is left to rely upon past experience in dealing with the speaker to perceive attributes of the speaker's voice that indicate emotional states. For example, from past interactions, the listener may be able to determine that the speaker is stressed, happy, or sad. That is, the listener must rely upon hearing and judgment to gauge the speaker's state of being.

[0003]   The ability to determine such attributes can be beneficial, particularly in cases where one has had no previous interaction, or minimal interaction, with a call participant. In such cases, recognizing the call participant's emotional state would be difficult, if not impossible. For example, in cases where one is engaging in a conference call with unknown individuals, the ability to determine another's emotional state, or biometric information that provides an indication of one's state of being, can be beneficial.

[0004]   The ability to determine one's state of being can be beneficial in other situations such as checking on a child. Children tend to be less talkative when in the company of adults. For example, when checking on a child that is in the presence of others, such as a babysitter, the child is likely to respond to inquiries about the child's well being with uninformative answers such as "I'm fine" or "I'm okay". The ability to determine the child's state of being would provide parents with valuable information as to the child's welfare.

SUMMARY OF THE INVENTION

[0005]   The present invention provides a method, system, and apparatus for determining biometric information from a speaker's voice.   More particularly, in accordance with the inventive arrangements disclosed herein, biometric information such as a speaker's voice level, stress level, voice inflection, and emotion can be determined from voice signals received over a telephone call. The biometric information can be encoded and provided to a subscriber engaged in the call with the speaker.

[0006]   One aspect of the present invention can include a method of providing biometric information over an established telephone call between a speaker and a subscriber. The method can include receiving voice information from the speaker over the call, determining biometric information from the voice information of the speaker, encoding the biometric information, and sending the biometric information to the subscriber over the call. That is, the biometric information can be sent as an encoded stream of information embedded within the voice stream of the call.

[0007]   The determining step can include extracting at least one attribute from the voice information, comparing the at least one attribute with voice metrics, and generating the biometric information based upon the comparing step. The encoding step can include removing inaudible portions of the voice information and embedding the biometric information in place of the inaudible portions within a voice stream carried over the call.

[0008]   The biometric information can specify an indication of a speaker's voice level, stress level, voice inflection, and/or an emotional state. Notably, the subscriber can receive the biometric information and voice signals, both of the speaker, substantially concurrently over the call. The method further can include decoding the received biometric information and presenting the information to the subscriber.

[0009]   Other embodiments of the present invention can include a system having means for performing the various steps disclosed herein and a machine readable storage for causing a machine to perform the steps described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010]   There are shown in the drawings, embodiments which are presently preferred, it being understood, however, that the invention is not limited to the precise arrangements and instrumentalities shown.

[0011]   FIG. 1 is a schematic diagram illustrating a system for determining biometric information from a speaker's voice in accordance with one embodiment of the present invention.

[0012]   FIG. 2 is a flow chart illustrating a method of determining biometric information from a speaker's voice in accordance with another embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0013]   FIG. 1 is a schematic diagram illustrating a system 100 for determining biometric information from a speaker's voice in accordance with one embodiment of the present invention. As shown, the system 100 can include a voice analysis system 105 which can be communicatively linked with a communications network 130 such as the Public Switched Telephone Network (PSTN).

[0014]   The voice analysis system 105 can be implemented in an information processing system, such as a computer system or server having a telephony interface, for example one communicatively linked with a telephone switching system, or a processing card disposed within a telephone switching system. As such, the voice analysis system 105 can be patched into a telephone call between two or more callers. For example, a subscriber to a service involving the voice analysis system 105 can invoke the service using one or more touch tone keys prior to a call or during an ongoing call.

[0015]   The voice analysis system 105 can include a biometric analysis engine 110, a comparator 115, a data store 120 including voice metrics, and a biometric information encoder 125. The biometric analysis engine 110 can extract biometric information from speech or voice signals received over a telephone call. For example, the biometric analysis engine 110 can determine one or more attributes that are indicative of a speaker's voice level, stress level, voice inflection, and/or emotional state.

[0016]   The comparator 110 can compare the attributes determined by the biometric analysis engine 110 with one or more voice metrics stored in the data store 120. The voice metrics can be a collection of empirically determined attributes representing various voice levels, stress levels, voice inflections, and/or emotional states. Based upon the comparison of speaker voice attributes with voice metrics, the comparator 110 can generate biometric information which can specify an indication of a speaker's voice level, stress level, voice inflection, and/or emotional state.

[0017]   The biometric information encoder 125 can encode any generated biometric information for transmission to the subscriber. Once the biometric information is generated, the information can be encoded and provided as embedded digital information within a digital voice stream of a telephone call. As such, one aspect of the

biometric information encoder 125 can be implemented as a perceptual audio processor, similar to a perceptual codec, to analyze a received voice signal. A perceptual codec is a mathematical description of the limitations of the human auditory system and, therefore, human auditory perception. Examples of perceptual codecs can include, but are not limited to MPEG Layer-3 codecs and MPEG Layer-4 codecs. The biometric information encoder 125 is substantially similar to the perceptual codec with the noted exception that the biometric information encoder 125 can, but need not implement, a second stage of compression as is typical with perceptual codecs.

[0018] The biometric information encoder 125, similar to a perceptual codec, can include a psychoacoustic model to which source material, in this case a voice signal from a call participant, can be compared. By comparing the voice signal with the stored psychoacoustic model, the perceptual codec identifies portions of the voice signal that are not likely, or are less likely to be perceived by a listener. These portions are referred to as being inaudible. Typically a perceptual codec removes such portions of the source material prior to encoding, as can the biometric information encoder 125.

[0019] Still, those skilled in the art will recognize that the present invention can utilize any suitable means or techniques for encoding biometric information and embedding such digital information within a digital voice stream. As such, the present invention is not limited to the use of one particular encoding scheme.

[0020] In operation, a telephone call can be established between a subscriber 135 and a speaker 140 over the communications network 130. The speech or voice signals of the speaker 140 can be provided to the voice analysis system 105. The voice analysis system 105 can determine biometric information from the speaker's 140 voice. The biometric information can be provided to the subscriber 135 as embedded digital information within a digital voice stream of the telephone call.

[0021] It should be appreciated that the subscriber 135 can be equipped with a suitable telephony device that is capable of decoding the received biometric information and presenting the information to the subscriber 135. For example, in one embodiment, the information can be displayed upon a display incorporated within or attached to the subscriber's 135 telephony device. In another embodiment, the biometric information can be presented audibly. That is, the information can be decoded and provided to a

text to speech system to be played to the subscriber through the subscriber's telephony device or another device communicatively linked to the telephony device.

[0022] FIG. 2 is a flow chart illustrating a method 200 of determining biometric information from a speaker's voice in accordance with another embodiment of the present invention. The method can begin in step 205 where a user subscribes to a voice analysis service. In step 210, the subscriber can establish a telephone call with another person, a speaker.

[0023] In step 215, the subscriber can invoke the voice analysis service. As noted, the subscriber can invoke the service by keying a particular code or sequence of digits or by issuing a spoken request. In step 220, voice signals from the speaker, that is the other call participant, can be provided to the voice analysis system. For example, the voice analysis system can be patched into the telephone call such that the voice signals from the speaker are provided to the voice analysis system for processing. In step 225, the voice analysis system determines one or more attributes of the voice signal.

[0024] In step 230, the voice analysis system compares the attributes with defined voice metrics. In step 235, the voice analysis system determines biometric information relating to the speaker's voice. As noted, the biometric information can include, but is not limited to, a speaker's voice level, stress level, voice inflection, and/or emotional state. The biometric information, for example a rating of voice level, stress level, voice inflection, and/or indication of an emotional state including a measure of degree of the emotional state, can be determined through the comparison of the speaker's voice attributes with the established voice metrics.

[0025] In step 240, the voice analysis system sends the biometric information to the subscriber over the established telephone call. More particularly, the biometric information can be sent to the subscriber as an encoded stream of digital information that is embedded within the digital voice stream. The biometric information encoder can identify which portions of the received audio signal are inaudible, for example using a psychoacoustic model.

[0026] For instance, humans tend to have sensitive hearing between approximately 2 kHz and 4 kHz. The human voice occupies the frequency range of approximately 500 Hz to 2 kHz. As such, the biometric information encoder can remove portions of a voice

signal, for example those portions below approximately 500 Hz and above approximately 2 kHz, without rendering the resulting voice signal unintelligible. This leaves sufficient bandwidth within a telephony signal within which the biometric information can be encoded and sent within the digital voice stream.

[0027] The biometric information encoder further can detect sounds that are effectively masked or made inaudible by other sounds. For example, the biometric information encoder can identify cases of auditory masking where portions of the voice signal are masked by other portions of the voice signal as a result of perceived loudness, and/or temporal masking where portions of the voice signal are masked due to the timing of sounds within the voice signal.

[0028] It should be appreciated that as determinations regarding which portions of a voice signal are inaudible are based upon a psychoacoustic model, some users will be able to detect a difference should those portions be removed from the voice signal. In any case, inaudible portions of a signal can include those portions of a voice signal as determined from the biometric information encoder that, if removed, will not render the voice signal unintelligible or prevent a listener from understanding the content of the voice signal. Accordingly, the various frequency ranges disclosed herein are offered as examples only and are not intended as a limitation of the present invention.

[0029] The biometric information encoder can remove the identified portions, i.e. those identified as inaudible, from the voice signal and add the biometric information in place of the removed portions of the voice signal. That is, the biometric information encoder replaces the inaudible portions of the voice signal with digital biometric information. As noted, the biometric information can include, but is not limited to, voice levels, stress levels, voice inflections, and/or emotional states as may be determined from a speaker's voice.

[0030] As noted, the biometric information can be decoded in the subscriber's telephone device, or a device attached to the subscriber's telephony equipment. The biometric information then can be present to the subscriber in a visual format or played through an audio interface. Notably, the biometric information can be received with the speaker's voice stream such that the subscriber is able to be presented with the

biometric information of the speaker while engaged in the telephone call and hearing the speaker's voice.

[0031] The method 200 has been provided for purposes of illustration only. As such, it should be appreciated that one or more of the steps disclosed herein can be performed in differing order depending upon the particular configuration of the present invention. For example, the subscriber can invoke the voice analysis service at any time prior to a call or during a call. Additionally, the present invention can be used regardless of whether the subscriber initiates a call or receives a call.

[0032] Further, in one embodiment, the subscriber can specify which voice stream is to be analyzed, for example by keying in a telephone number of the party or voice source. Such an embodiment can be useful in the context of conference calls. In any case, the examples disclosed herein are not intended as a limitation of the present invention.

[0033] The present invention can be realized in hardware, software, or a combination of hardware and software. The present invention can be realized in a centralized fashion in one computer system, or in a distributed fashion where different elements are spread across several interconnected computer systems. Any kind of computer system or other apparatus adapted for carrying out the methods described herein is suited. A typical combination of hardware and software can be a general purpose computer system with a computer program that, when being loaded and executed, controls the computer system such that it carries out the methods described herein.

[0034] The present invention also can be embedded in a computer program product, which comprises all the features enabling the implementation of the methods described herein, and which when loaded in a computer system is able to carry out these methods. Computer program in the present context means any expression, in any language, code or notation, of a set of instructions intended to cause a system having an information processing capability to perform a particular function either directly or after either or both of the following: a) conversion to another language, code or notation; b) reproduction in a different material form.

[0035] This invention can be embodied in other forms without departing from the spirit or essential attributes thereof. Accordingly, reference should be made to the

following claims, rather than to the foregoing specification, as indicating the scope of the invention.